Journal of Chromatography, 544 (1991) 393–411
Elsevier Science Publishers B.V., Amsterdam

CHROM. 23 187

Factor analysis and experiment design in high-performance liquid chromatography

X. Chemometric characterization of packings, solvents and solutes with hierarchical ascending classification and correspondence factor analysis[a]

MICHEL RIGHEZZA
Laboratoire de Chimiométrie, Université d'Orléans, B.P. 6759, 45067 Orléans Cedex 2 (France)
and
JACQUES R. CHRÉTIEN*
Laboratoire de Chimiométrie, Université d'Orléans, B.P. 6759, 45067 Orléans Cedex 2 and Institut de Topologie et de Dynamique des Systèmes, Associé au CNRS UA 34, 1 Rue Guy de la Brosse, 75005 Paris (France)

ABSTRACT

A method for the characterization of chromatographic packings and systems for liquid chromatography is presented. A model series of 36 chalcones was studied on 43 chromatographic systems corresponding to 31 different packings of various polarities. The behaviour of these compounds, in the space of the chromatographic systems, was analysed systematically with chemometric methods. The relative distances between packings used with different eluents, or between solutes, were studied using hierarchical ascending classification and correspondence factor analysis. These data processing techniques offer synthetic views of relative selectivities of packings. An in-depth analysis of these selectivities requires the study, with the same attention, of the role of the packings, of the eluents and of the solutes.

INTRODUCTION

The study of chromatographic processes at the highest level presupposes that equal attention is devoted to packings and solutes simultaneously. The characterization of packings and the modelling of solutes behaviour are strongly interdependent. Packings are characterized with selected solutes and the behaviour of solutes is modelled with reference to packings. The strong link between packings and solutes requires the use of chemometric methods to analyse these chromatographic processes [1].

This paper demonstrates how one can obtain general or specific information dealing with the characterization of chromatographic packings, solvents and solutes. The problem that arises is how to compare the behaviours of stationary phases,

[a] For Part IX see *Chromatographia*, 29 (1990) 259–264.

solvents and solutes. Owing to the wide range of types of interactions between solutes, packings and solvents, it is impossible to give a simple and general answer to such a question. Specific scales of polarity or hydrophobicity, related to specific types of packings, have been developed to classify solvents [2–5]. Linear or exponential equations can give good correlations with the expected behaviour of solutes [6–9]. However, with such relative scales, the number of proposed classifications and/or equations becomes proportional to the number of packing types and/or solutes series.

As opposed to developing many specific scales, the exploitation of the concepts of distances in a hyperspace offers more general approaches. It can be the exploitation of the distances between compounds in the hyperspace defined by the chromatographic systems or the analysis of the distances between these chromatographic systems in the hyperspace defined by the solutes. In this paper, the information content of a large set of 1548 chromatographic data, corresponding to a matrix giving the behaviour of a homogeneous series of 36 chalcones studied on 43 different chromatographic systems, will be presented by using appropriate dataprocessing techniques. This information content is centred on an evaluation of the relative performances of chromatographic packings, solvent effects and solute behaviour with the help of hierarchical ascending classification (HAC) and correspondence factor analysis (CFA).

EXPERIMENTAL

Data have been published in previous papers in this series [10–17]. Sixty-three compounds were eluted with 43 chromatographic systems.

This data set covers a large experimental field of chromatographic systems. As shown in Fig. 1, a three-dimensional space helps to represent the chromatographic experiments. The x-axis is associated with normal-phase high-performance liquid chromatography (NP-HPLC) and the y-axis with the reversed-phase mode (RP-HPLC). The main eluents are polar solvents such as water in the RP-HPLC mode or non-polar solvents such as heptane in the NP-HPLC mode. On the $0x$, $0y$ plane, solid lines indicate the nature of the solvent used as a binary mixture. Non-miscible solvents represented by dotted lines correspond to impossible experiments. Non-studied systems are represented by dashed lines. The z-axis is related to the packing of the column, without reference to any polarity scale.

The main characteristics of the data matrix are shown in Fig. 2a and b. Fig. 2a gives the paper number and its reference for each subset of chromatographic data. The composition of the solvent is indicated: heptane plus a given modifier or co-solvent x (3 or 0.5% by volume), or a water–methanol mixture (30:70, v/v). The commonly used labellings of the column packing are given with their corresponding identifiers used in the data processing. Five packings used in the normal-phase mode are labelled AP, DNAP, DNB, TCP and TB, which represent the bonded phases *n*-propyl picryl ether, 3-(2,3-dinitroanilino)propyl, 3,5-dinitrobenzamidopropyl, tetrachlorophthalimidopropyl and caffeine, respectively [17].

The chromatographic data were obtained with 30 different normal- and reversed-phase HPLC columns, in conjunction with 11 aqueous or non-aqueous eluents. These 30 HPLC columns packed with polar or non-polar packings and the different eluents used give rise to 43 chromatographic systems. Further, the 63 compounds and the 43 chromatographic systems define an incomplete data matrix of

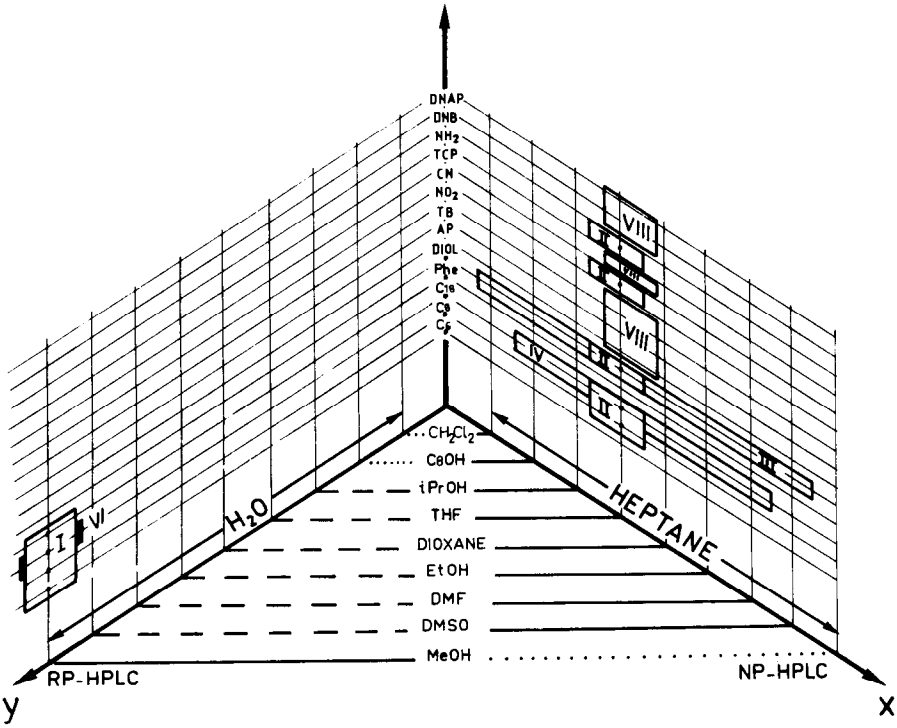
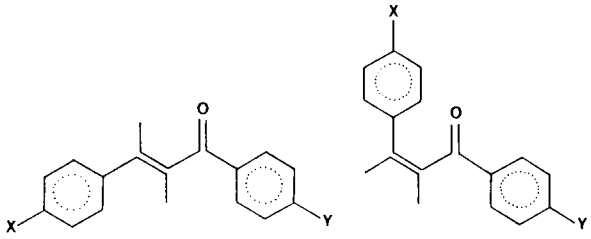


Fig. 1. Experimental field covered by a systematic investigation in HPLC. The behaviour of a large series of 53 chalcones ($XC_6H_5CH = CHCOC_6H_5Y$) and sixteen test compounds was studied with 43 normal- and reversed-phase HPLC systems. The x-axis refers to the NP mode and the y-axis refers to the RP mode. The z-axis indicates the packings without reference to any polarity scale. The dotted lines indicate non-miscible solvents. The dashed lines imply experiments that were not performed. Here and elsewhere, iPr = isopropyl, Et = ethyl; Me = methyl; C_8OH = octanol; THF = tetrahydrofuran; DMF = dimethylformamide; DMSO = dimethylsulphoxide.

capacity factors (k'). The original k' data matrix can be divided into two sets, the studied matrix, which is limited to the 36 chalcones studied with the 43 chromatographic systems indicated by bold lines in Fig. 2a, and the complementary raw data. We shall describe the treatment of these complementary data in a forthcoming paper.

The compounds labelled with their identifier in Fig. 2a are detailed in Fig. 2b. Fifty-three compounds were substituted chalcones. The general formulae of the chalcones considered are



where X and Y are as listed in Fig. 2b. One of the main interests of this series is to offer a large variation of the $+I$ and $-I$ electronic effects due to the aromatic substituents. The (*E*)-*S*-*cis*-chalcones are denoted in the text as X–Y and the corresponding (*Z*)-*S*-*cis* isomers are denoted with an asterisk, X–Y*. Sixteen compounds most often proposed as test compounds for testing chromatographic systems were used.

DATA PROCESSING

The data matrix was studied first with HAC to analyse the main trends of the chromatographic systems and solutes. HAC analysis of data matrix subsets gives the particular behaviour of specific systems. HAC allows distances between points in a multi-dimensional space to be analysed by progressive grouping and the successive two closest points to be represented in the form of a classification tree. To do this, the data points are gathered in n subclasses by testing their relative proximity. Distances between objects have been taken as the chi-squared distance [15]. The successive classes are grouped in the same manner in $n-1$, $n-2$, $n-3$, ..., classes until a single class is obtained. The aggregation level is given by the y-axis with a relative scale

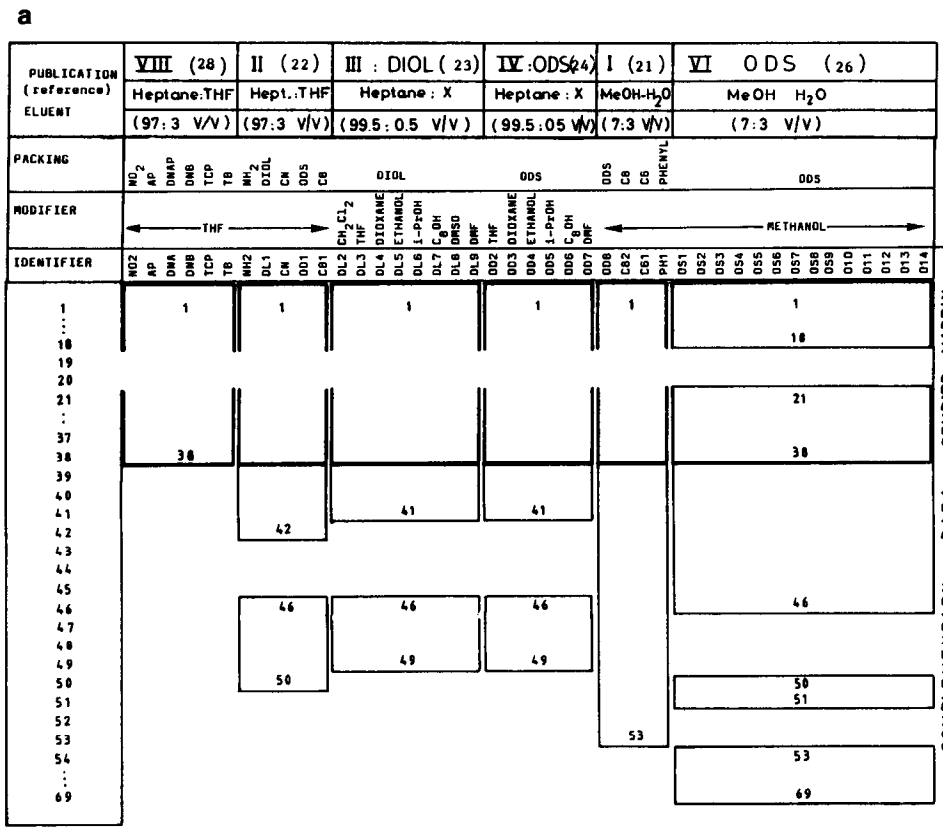


Fig. 2.

b

CHALCONES

	X - Y		X - Y		X - Y
1	H - CF_3	24	H - H*	47	Me - Phenyl*
2	H - tBu	25	F - H*	48	MeO - Phenyl*
3	H - iPr	26	H - F*	49	MeO - NO_2*
4	H - H	27	H - Et*	50	NO_2 - MeO*
5	F - H	28	H - Me*	51	NO_2 - NO_2*
6	H - F	29	F - Me*	52	NH_2 - H*
7	H - Et	30	F - F*	53	H - OH*
8	H - Me	31	MeO - Me*		
9	F - Me	32	Me - MeO*		
10	F - F	33	F - MeO*		
11	MeO - Me	34	H - NO_2*		
12	Me - MeO	35	NO_2 - Me*		
13	F - MeO	36	NO_2 - H*		
14	H - NO_2	37	MeO - MeO*		
15	NO_2 - Me	38	NO_2 - F*		
16	NO_2 - H	39	F - NO_2		
17	MeO - MeO	40	MeO - NO_2		
18	NO_2 - F	41	$N(Me)_2$ - NO_2		
19	MeO -Phenyl	42	NO_2 - MeO		
20	Me -Phenyl	43	NO_2 - NO_2		
21	H - CF_3*	44	NH_2 - H		
22	H - tBu*	45	H - OH		
23	H - iPr*	46	F - NO_2*		

Test compounds

54	Nitrobenzene
55	Naphthalene
56	Phenanthrene
57	Methyl benzoate
58	Biphenyl
59	Diethyl phthalate
60	Anthracene
61	4-methyl-cresol
62	2-phenyl-ethanol
63	Benzophenone
64	Benzyl alcohol
65	3-phenyl-propanol
66	4-phenyl-butanol
67	6-phenyl-hexanol
68	9-phenyl-nonanol
69	1-nitro-naphthalene

Fig. 2. (a) Data matrix of over 2000 capacity factors of 69 selected compounds analysed with 43 different chromatographic systems. A homogeneous set of 36 × 43 data was selected for data processing as indicated by bold lines. Remaining subsets of chromatographic data correspond to the complementary data indicated. (b) Description of the model compounds: chalcone ($XC_6H_5CH = COC_6H_5Y$), (*E*)-*S-cis* (X–Y), (*Z*)-*S-cis* (X–Y*) and test compounds most often used in HPLC. tBu = *tert.*-butyl.

varying from 0 to 100%. The x-axis represents only a qualitative ordering according to the progressive increase in the aggregation criterion of the main subsets.

Another way to study the data matrix is by CFA. This statistical method aims to search for the minimum number of factors and eigenvectors which the raw data matrix is able to reproduce. The first axis corresponds to the maximum of inertia of the cloud of data points. Other axes of this space are calculated perpendicular to the first and to each other. Species (solutes) and variables (chromatographic systems) can be drawn into the multi-dimensional space defined or projected onto planes. The first axis corresponds to the main trend; it often represents more than 80% of the information content in HPLC [10–17]. It can easily be related to physico-chemical parameters [10,12]. The physico-chemical explanation of the other axes becomes more difficult as the information content of the axis decreases. This is due to the mathematical definition of the method; only the first axis is strictly determined by the

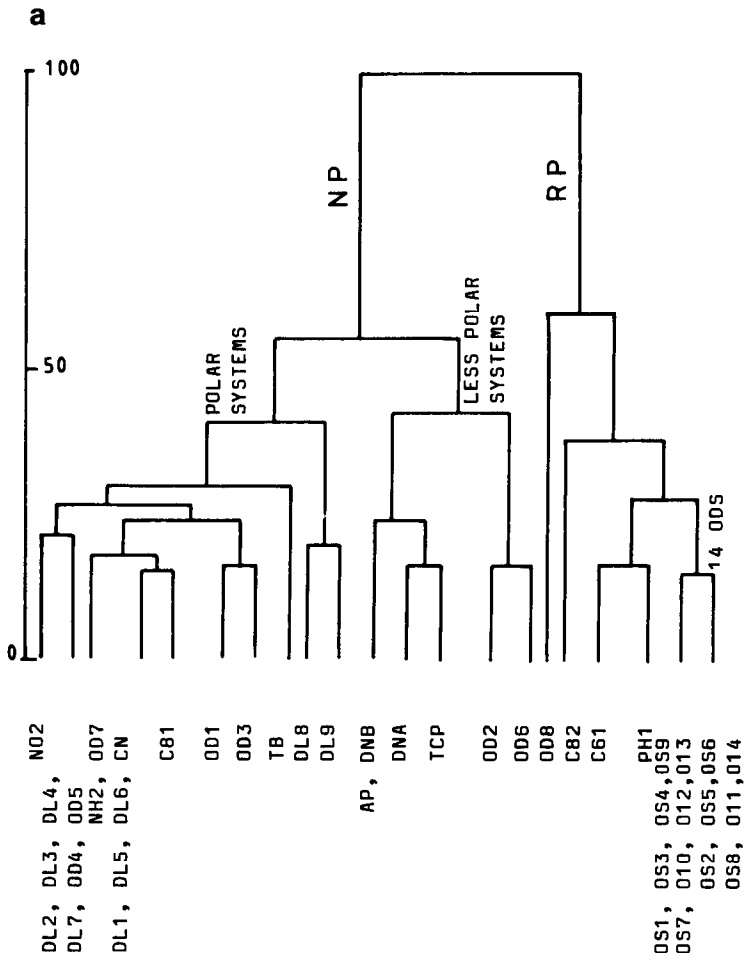


Fig. 3.

maximum of inertia of the raw data matrix, whereas the other axes are conventionally chosen perpendicular to the preceding ones.

HAC and CFA were used step by step in the same way in order to exploit their complementarity.

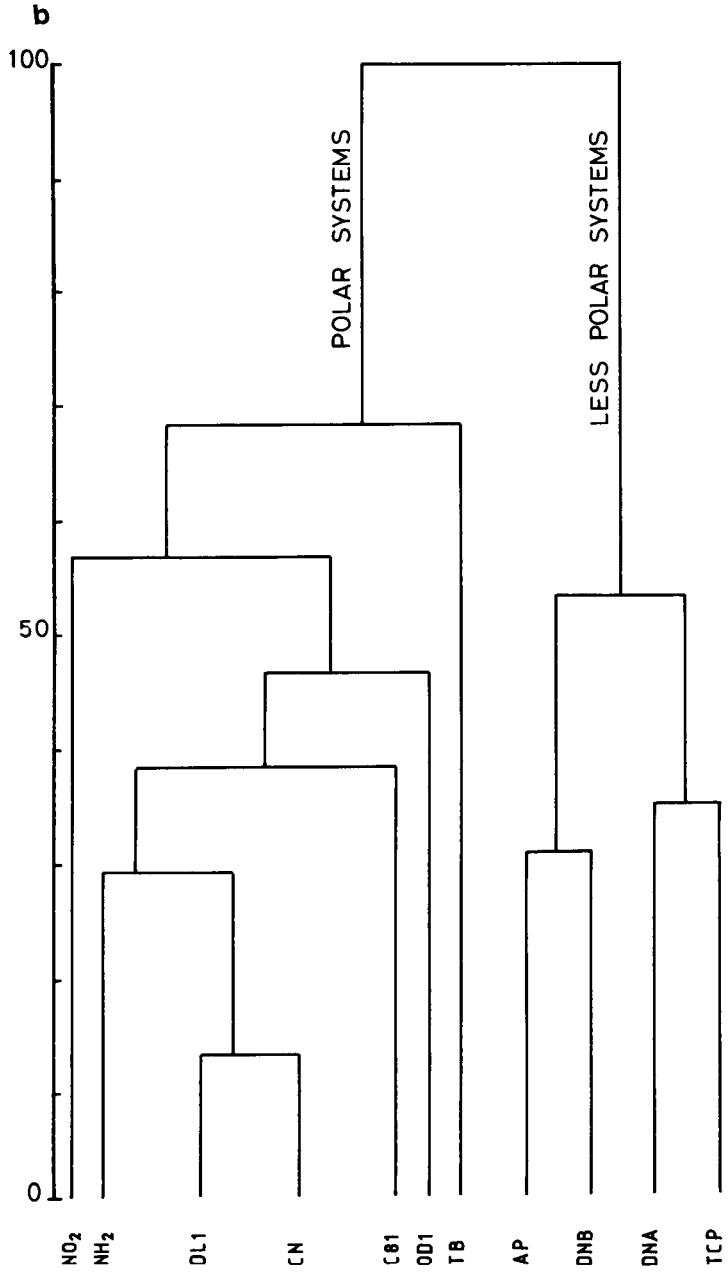


Fig. 3.

(Continued on p. 400)

RESULTS AND DISCUSSION

HAC and CFA statistical methods were applied to the complete data matrix to study the chromatographic systems and solute behaviour. Subsets of this matrix were defined to improve the study of specific chromatographic systems or the structural effects on the retention.

HAC

HAC of the complete data matrix makes it possible to classify chromatographic systems (Fig. 3a) in the space defined by compounds, *i.e.*, by taking compounds as a probe to reveal analogies or differences in the chromatographic systems. Two main groups are extracted: normal-phase systems (NPS) and reversed-phase systems (RPS). We shall use here non-classical NPS and RPS terminology to distinguish the well known abbreviations NP and RP most often given to the packings in relation to the chromatographic mode used. This terminology takes into account the fact that different packings were used in the two modes depending on the nature of the eluent.

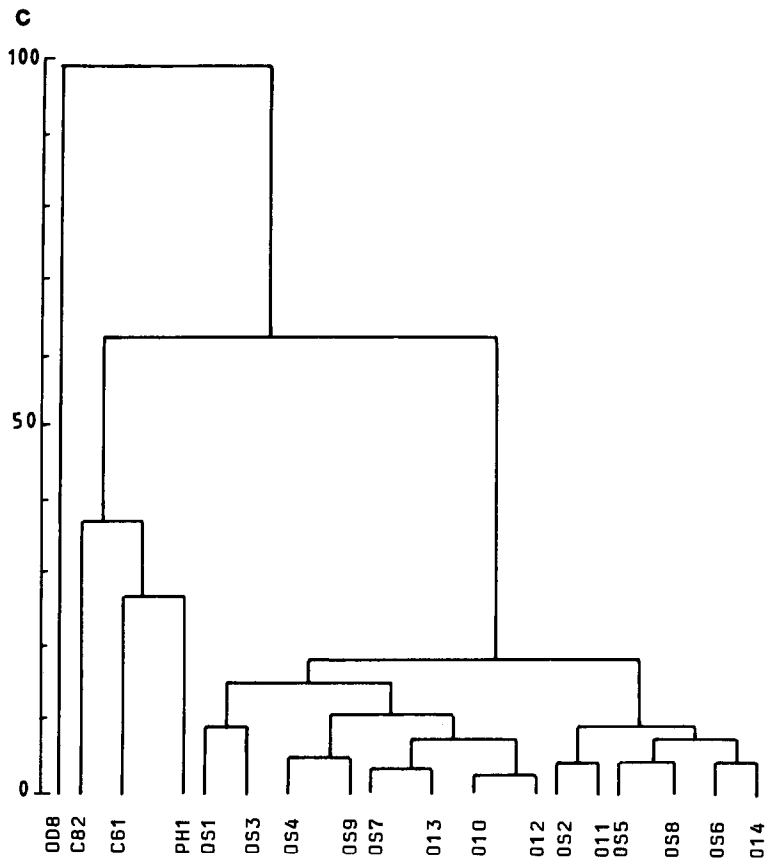


Fig. 3.

NPS contains polar and non-polar chromatographic systems such as NO_2, diol and ODS bonded phases used in conjunction with non-aqueous mobile phases. RPS contains a weakly polar bonded phase (phenyl) and non-polar bonded phases (C_8, ODS) used with aqueous solvents (methanol–water). The level of aggregation for both class-

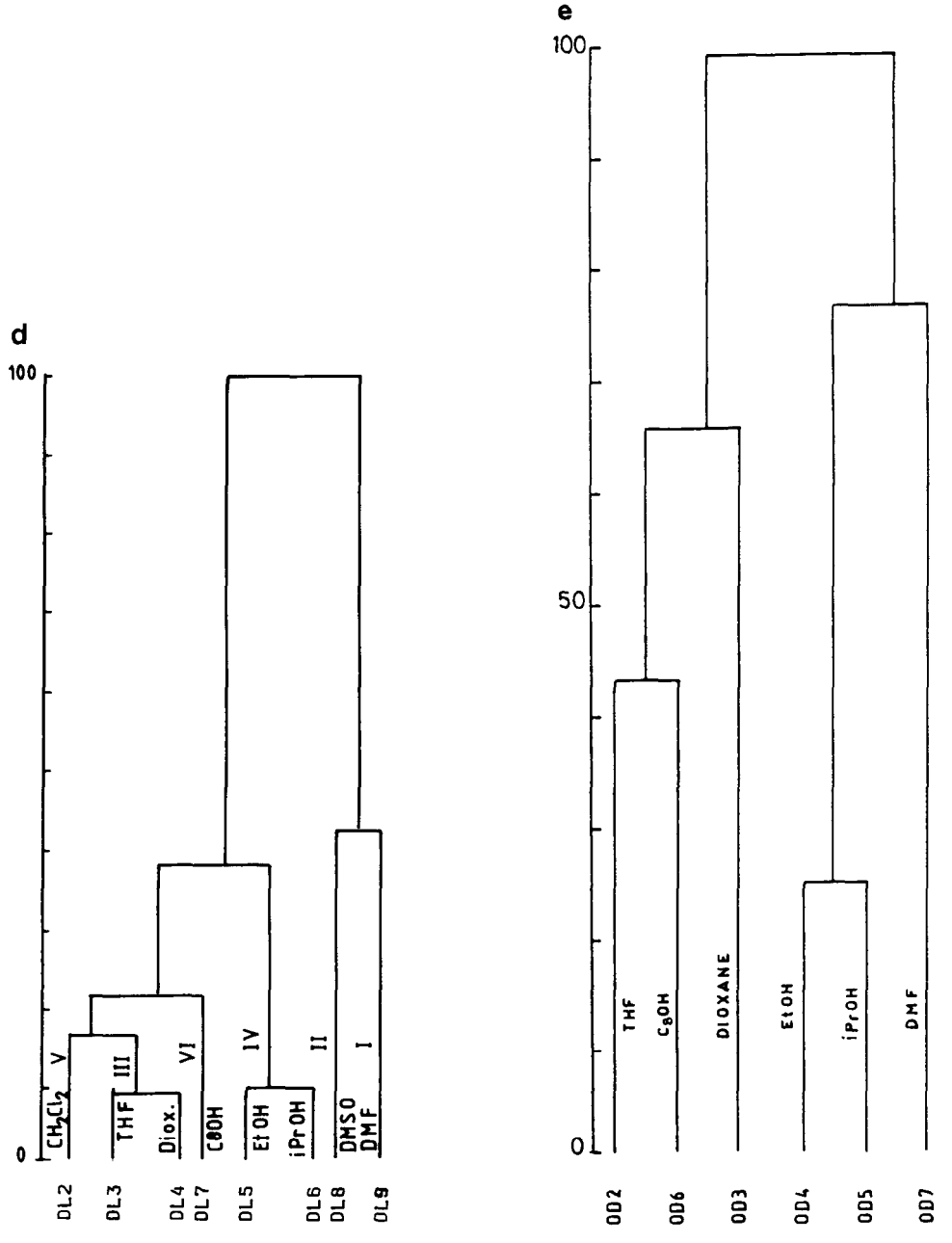


Fig. 3.

(Continued on p. 402)

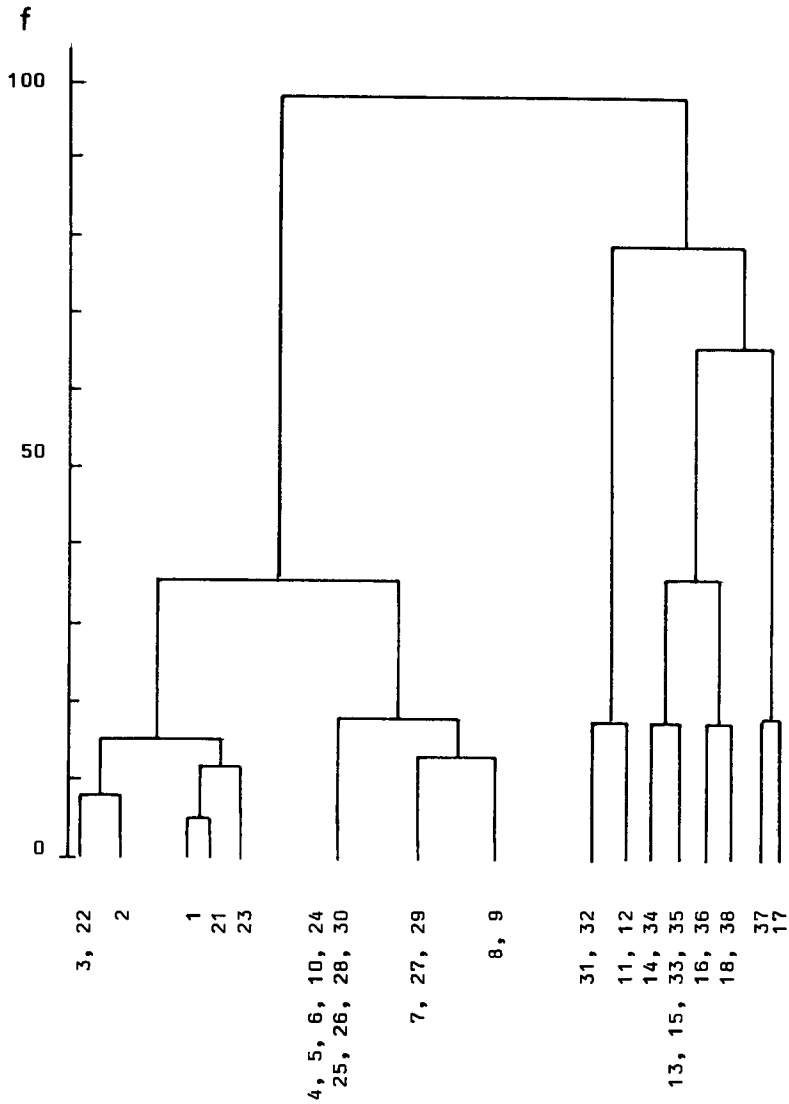


Fig. 3. Hierarchical ascending classification of (a) chromatographic systems; (b) normal-phase (NP) chromatographic systems; (c) reversed-phase (RP) chromatographic systems; (d) diol chromatographic systems; (e) ODS (RP-18) chromatographic systems; (f) solute with the complete data matrix.

es is 100%. Each group is divided into two subclasses with a level of aggregation of *ca.* 55%. The NPS group is divided into two classes. Chromatographic systems are aggregated with a criterion of relative similarity. The chi-squared distance does not reflect the polarity of systems but the selectivity. Using the chi-squared distance it is not surprising that the first group is composed principally of NO_2, NH_2 and diol phases and the second group AP, DNB, DNA and TCP systems. With these polar

systems two specific mechanisms of interaction are involved. The RPS group shows two classes constituted by one system (ODS) and seventeen other systems.

To classify the chromatographic packings, we selected two subset data matrices. For these data matrices, all experiments were performed with a constant nature and concentration of the eluent. With NPS, the eluent was heptane–THF (97:3, v/v) and with RPS the eluent was methanol–water (7:3, v/v). In both subset data matrices, the solvent effects were constant. The NPS subset (Fig. 3b), at a level of aggregation of 100%, is divided into two subgroups. In the first subgroup diol, CH_2, NH_2, C_8, ODS, NO_2 and TB are aggregated consecutively. The aggregation level reflects the differences between the phases. The diol, CN and NH_2 phases are relatively similar but different from NO_2. The second subgroup shows that AP and DNB phases and DNA and TCP phases present the same selectivity. The HAC of the RPS subset is given in Fig. 3c. At a level of aggregation of 100% an artefact is detected: the classification shows that the OD8 packing differs from the fourteen other ODS bonded phases [15]. This result cannot be explained. In this set of data the solvent effects are constant. The HAC of the four columns [10] shows that there is a good similarity between the fourteen ODS phases, which are clearly different from C_8, C_6 and phenyl bonded phases. These three phases are aggregated at an aggregation level of 38%.

The solvent effects can be shown simply. We have extracted from the original data matrix two subset data matrices with the same packing material. These two matrices were studied using diol (Fig. 3d) and ODS bonded phase (Fig. 3e). Both data matrices were created for a constant concentration of modifier; only the pure solvent effects are observed. Solvents are aggregated according to the behaviour of solutes which present different retentions. Some solutes have polar and others have non-polar substituents. The retention is directly influenced by the nature of these substituents, so the HAC related to solvents is characteristic of the system selectivity for solutes. The HAC of solvents taking into account solutes without any discrimination concerning their polarity should give classes depending on Snyder's classification [2]. Indeed, with a diol bonded phase (Fig. 3d), at the aggregation level of 100%, two classes are formed. DMF (I) and DMSO (II) are aggregated in the same class at an aggregation level of *ca.* 40%. The aggregation level is higher than for the other pairs of eluents considered. The corresponding selectivities of DMF and DMSO are different from the other modifiers. In the second class, at a 10% aggregation level two classes are found: THF and dioxane (III) and ethanol and isopropanol (IV). Dichloromethane (V) is aggregated with class III at 17% and octanol (VI) is aggregated at an aggregation level of 22%. Solvents are arranged in six successive classes. In the same manner, with the ODS bonded phase (Fig. 3e), a different eluent classification is obtained. At a level of 45%, four classes are determined: DMF (I), ethanol and isopropanol (II), dioxane (III) and THF and octanol (IV). The difference in classification of THF, octanol and dioxane with ODS and diol-bonded phases is due to specific phase selectivities. Classification of solvent selectivity is linked to the chromatographic mode and the nature of the bonded phase and of the solutes. A relative scale of solvent polarity could be obtained by selecting subset data matrices of solutes with the same type of interaction.

In the same way, solutes are classified according to chromatographic systems taken as references or probes (Fig. 3f). The main trend of HAC of the complete data matrix represents the affinity of the solutes for the chromatographic systems. HAC of

chromatographic systems delineates two groups, so it is not surprising that HAC of solutes also reveals two groups of compounds. At an aggregation level of 100%, solutes on the left have affinity for RPS and on the right they have affinity for NPS. In each group, (*E*)-*S*-*cis* and (*Z*)-*S*-*cis* isomers are closely related in the classification tree (solutes 1–21, 2–22, ...) and the effect of the compound configurations on the retention is weak. The RPS-dependent group is composed of the alkyls, H, F and CF_3 substituted isomers. Two subgroups of (*E*)-*S*-*cis* and (*Z*)-*S*-*cis* isomers, H–CF_3, H–tBu, H–iPr and H–H, H–F, F–H, H–Et, H–Me, F–Me and F–F have an aggregation level of 35%. The mechanism of retention is governed by hydrophobic interactions. The NPS-dependent group is composed of isomers which have a strong electronic effect due to donor–acceptor effects of the substituents such as MeO–Me, Me–MeO, H–NO_2, NO_2–H, NO_2–Me, MeO–MeO, NO_2–F and F–MeO. At an aggregation level of 80%, MeO–Me and Me–MeO form a subgroup. The complementary group is divided at an aggregation level of *ca.* 65% into two subgroups: MeO–MeO and F–MeO, H–NO_2, NO_2–Me, NO_2–H and NO_2–F. From the study of the NPS-dependent group it appears that the retention mechanism is governed by the electronic effects of substituents on delocalized π electrons in the aromatic structure.

Study of similarities by HAC offers a simple analysis of the complete data matrix. The observations and some explanations of the influences of chromatographic systems, solvents and substituent nature are given. It is possible to select, with an acceptable approximation, the best chromatographic system to separate solutes according to their respective affinity for the packings.

CFA

The projections of the chromatographic systems on planes defined by factorial axes 1 and 2 and axes 2 and 3 are given in Fig. 4a and b. The three best axes of inertia represent 84%, 10% and 2%, respectively, of the information content. In analogy with the HAC study, chromatographic systems and solute behaviour are examined separately. Selections of submatrices are made in order to expand the understanding of some specific chromatographic systems and solutes.

Because the complete data matrix covers a large domain of chromatographic modes, it is not surprising to see the chromatographic mode described by axis 1 which represents the main trend. Negative values on the abscissa correspond here to the normal-phase (NP) mode and positive values to the reversed-phase (RP) mode. On the NP side are found packing materials such as NH_2, NO_2, diol and also C_8 and ODS used with non-aqueous solvents. On the RP side, aqueous solvents were employed with phenyl and alkyl bonded phases.

Conventionally axes 2 and 3 are chosen perpendicular to the first axis. The second axis represents the contributions of the packings and the solvent nature to the specific interactions in the NP or RP mode. This is specially clear for the NP mode because different packing materials and solvents were used. The contribution of chromatographic systems can be related to the HAC results. From positive to negative values on second axis, similar systems are located in two groups: DNA, TCP, DNB, OD2 and OD6 with positive values on axis 2 compose the first group and NH_2, CN, NO_2 and diol with negative values on axis 2 compose the second group. This observation agrees with the HAC results.

The solvent contribution can be elucidated with the help of Table I. We have

classified modifiers into four groups, aprotic–non-basic (A), aprotic–weakly basic (B), aprotic–basic (C) and protic (D) solvents. As the contribution of the diol or ODS phase is kept constant, solvent effects appear clearly as a combination of three factors: concentration, the protic or aprotic group and dielectric constant. The DL3, DL1 and OD2, OD1 chromatographic systems show concentration effects on the solvent polarity. The concentration of the modifier increases from 0.5% to 3% of THF the polarity of the mobile phase increases simultaneously. The systems DL3 and

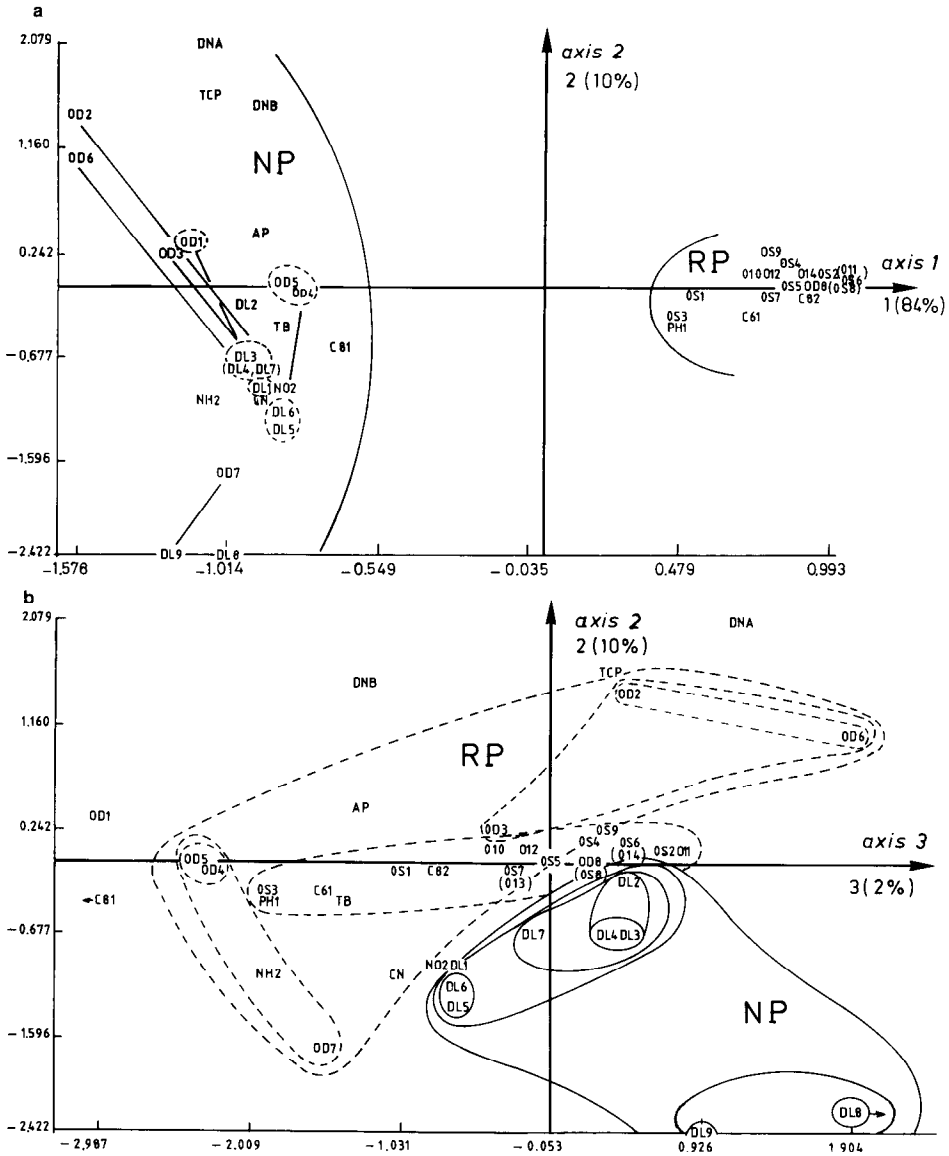


Fig. 4.

(Continued on p. 406)

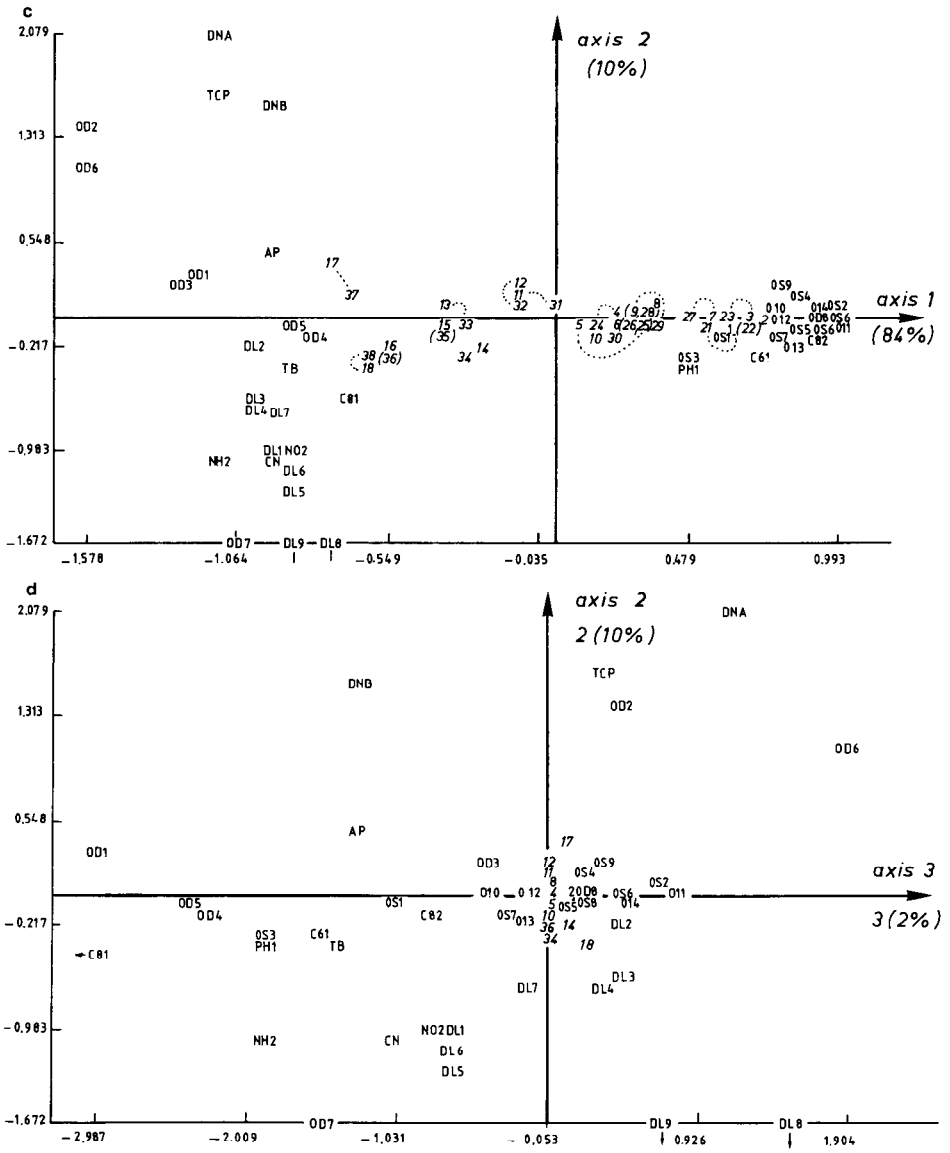


Fig. 4. Correspondence factor analysis. (a) Projection of chromatographic systems on the plane defined by the first and second best axes of inertia which represent 84% and 10%, respectively, of the information content. (b) Projection of chromatographic systems on the plane defined by the second and third best axes of inertia which represent 10% and 2%, respectively, of the information content. Straight lines join ODS and diol systems with the same eluent. Dashed lines delineate systems which have an identical projection. (c) Projection of compounds on the plane defined by the first and second best axes of inertia which represent 84% and 10%, respectively, of the information content. (d) Projection of compounds on the plane defined by the second and third best axes of inertia which represent 10% and 2%, respectively, of the information content.

TABLE I

PROTIC AND DIELECTRIC CHARACTERISTICS OF SOLVENTS

Chromatographic system[a]	Solvent (heptane + X)	Group[b]	Dielectric constant, ε
DL1	THF (3%)	B	7
DL2	CH_2Cl_2 (0.5%)	A	9
DL3	THF (0.5%)	B	7
DL4	Dioxane (0.5%)	B	2.2
DL5	Ethanol (0.5%)	D	24.5
DL6	Isopropanol (0.5%)	D	18
DL7	Octanol (0.5%)	D/B	10.3
DL8	DMSO (0.5%)	C	46.7
DL9	DMF (0.5%)	C	36.7
OD1	THF (3%)	B	7
OD2	THF (0.5%)	B	7
OD3	Dioxane (0.5%)	B	2.2
OD4	Ethanol (0.5%)	D	24.5
OD5	Isopropanol (0.5%)	D	18
OD6	Octanol (0.5%)	D/B	10.3
OD7	DMF (0.5%)	C	36.7

[a] DL*n* = diol bonded phase; OD*n* = ODS bonded phase.

[b] Groups: A = non-protic, non-basic; B = non-protic, weakly basic; C = non-protic, basic; D = protic.

OD2 are shifted to lower values on axis 2. When the concentration of the modifier is kept constant (0.5%), the coordinates of diol and ODS systems vary from positive to negative values with group, from group A to group D, and dielectric constants, from the lowest to the highest values. With both phases (DL*n*, OD*n*), solvents are arranged in four groups according to Table I.

Information relative to the different polarities of both diol and ODS phases can be obtained in this study. The solvent effects are similar for both phases but the polarity of the resulting chromatographic system defined by the packing material and the solvent is shifted from a negative value with the diol phase to a positive value with the ODS phase. Solvent effects and phase polarity show the same trend and the resulting contribution to the polarity of the chromatographic system is the sum of the individual contributions.

The third axis represents only 2% of the information content. As the first axis (84%) is related to the chromatographic mode and the second axis (10%) to the solvent effect, we have selected the projection on the plane defined by the second and third best axes of inertia (Fig. 4b). To explain the meaning of the third axis, the behaviour of two chromatographic systems with diol and ODS bonded phases is observed using results from the HAC study. Previous aggregations are drawn as solid and dotted lines for the diol and ODS bonded phases, respectively. Fig. 4b gives a good idea of the aggregation rules; all groups and subgroups delineated by HAC are present. HAC was explained with the help of solvent selectivity. With CFA, the third axis is linked to the selectivity of the solvent. Differences in solvent selectivity can be measured by projecting points onto this third axis.

Let us compare solvent effects with the two bonded phases. THF (DL3, OD2) is arbitrarily chosen as a reference; it has the same positive ordinates on the third axis in both instances. Octanol has a negative value with the diol bonded phase (DL7) and a positive value with the ODS bonded phase (OD6). DMF has positive (DL9) and negative (OD7) ordinates, respectively. Ethanol, isopropanol and dioxane are at the left-hand side of THF with the same order in both instances (DL5, DL6, DL4 and OD4, OD5, OD3). Two important inversions are observed with octanol and DMF. Considering that the total selectivity results of three major interactions (dipole–dipole, proton acceptor and proton donor [2]), it is clear that the weight of individual interactions is different from diol to ODS bonded phases. The chromatographic systems chosen plus the nature of the selected solutes define a reference system for solvent classification in terms of polarity and selectivity. This classification is to be performed again when the chromatographic system, the chromatographic mode, the type of bonded phase and/or the nature of solutes are modified.

Let us consider now the projection of solutes in the two planes defined by axes 1, 2 and 3 (Fig. 4c and d). The information contents of axes 2 (10%) and 3 (2%) are weak compared with axis 1 (84%). Axis 1 can be related to the chromatographic mode and axis 2 to the solvent effects, as shown previously. These figures show that the retention of solutes is more influenced by the chromatographic mode. The solute affinity for a chromatographic mode is given by the abscissa of the projection on axis 1. For example, let us examine three solutes, labelled 2, 17 and 31. Solute 2 has a high k' value in the RP mode and a small k' value in the NP mode. Solute 2 has a greater affinity for RP systems than for NP systems. Its abscissa is around 0.70. Solute 17 has a greater k' value in the NP mode than in the RP mode. Its abscissa is equal to -0.75. The affinity of solute 17 is higher for NP systems. Solute 31 has an abscissa around zero. Capacity factors in the RP and NP modes are identical. The affinity of solute 31 is equivalent in both systems. Consequently, the projection of solutes on axis 1 represents the chromatographic behaviour of the solute with respect to a single chromatographic system or to a large set of chromatographic systems. The distance between solutes is a measure of the difference in their capacity factors, *i.e.*, a measure of the selectivity. The selectivity is linked to the mechanism of solute retention. The selected series of chalcones is well adapted to reveal the hydrophobic and electronic effects of the substituting groups and the isomer configuration. To understand the combined effects, subseries of (*E*)-*S*-*cis* isomers are selected, where X or Y is H. In Fig. 5a and b, we have drawn the projection of solutes on axis 1 for the complete data matrix and four subsets of the data matrix. The H–H solute is taken as the reference and the relative positions of NO_2, F, CF_3 and alkyl substituents are examined. (*E*)-*S*-*cis*- and (*Z*)-*S*-*cis*-chalcone isomers with X = H or Y = H have similar projections and chromatographic behaviour. This prompts us to examine only substituted (*E*)-*S*-*cis*-chalcones where X = H.

Obviously, projection of the complete data matrix does not make it possible to distinguish the hydrophobic and electronic effects (Fig. 5a). The CF3 substituent seems to have the same effect as an alkyl substituent and the projections of the Me and F substituents are too similar to explain the combined effects. Different subsets of packings have been studied by CFA. The packings are NO_2, CN and diol (Fig. 5b.1), diol (Fig. 5b.2) and ODS (Fig. 5b.3) in the NP mode and ODS (Fig. 5b.4) in the RP mode. Because of the subset data matrix selected, in these figures solutes which have

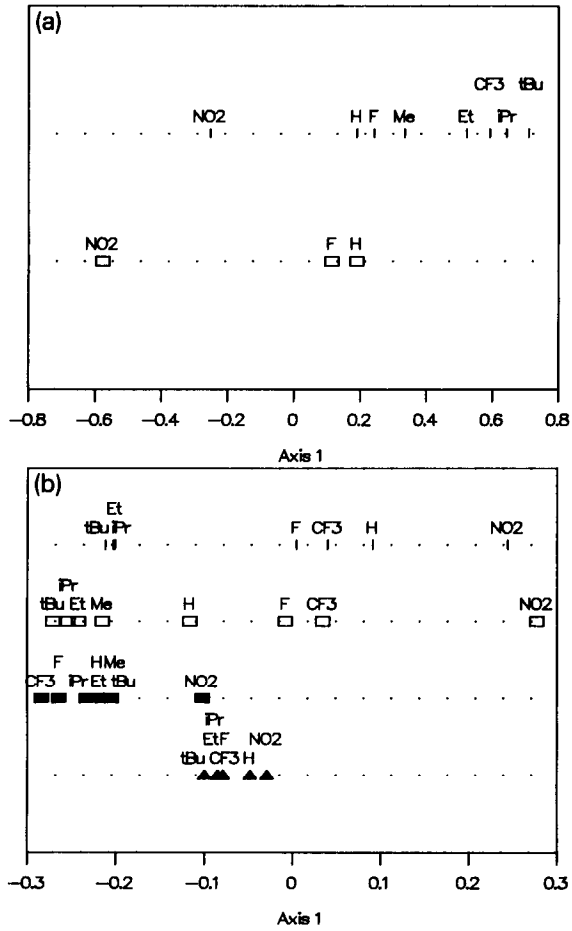


Fig. 5. Correspondence factor analysis of chromatographic behaviour of compounds: contribution of electronic effects or hydrophobicity on various chromatographic subsets. (a) Complete chromatographic system set with (*E*)-*S*-*cis*-chalcones with (1) X = H and (2) Y = H; (b) analysis of four chromatographic subsets: (1) NO_2, CN and diol, (2) diol, (3) ODS packings in the NP mode and (4) ODS packing in the RP mode.

affinity for NP systems are located on the right-hand side and solutes which present affinity for RP systems are on the left-hand side. With NP systems (Fig. 5b.1 and 2) electronic effects are preponderant. Alkyl substituents are far to the left, F, CF_3 and H are in the centre and NO_2 is far to the right. On the diol phase correlation was found between the capacity factor of phenylsubstituted isomers and Hammett constants [18]. We found here the same trend with chalcone isomers.

There are some differences between the two projections, owing to specific interactions in the case of NO_2 and CN phases. With the ODS phase in the NP mode, the projections are located on the left. There is no correlation with Hammett constants; the influence of the electronic effects is weak and the solute affinity for this NP chromatographic system is low. In the RP mode on the ODS phase (Fig. 5b.4),

solutes have a strong affinity with the stationary phase, and projections are located on the left near the centre. On this side of axis 1, as seen previously, the hydrophobic effect on the retention is preponderant.

CONCLUSION

To characterize packings, solvents and solutes systematically, the advantage of working directly with a large series of k' data is to avoid any *a priori* assumption about the chromatographic process. Obviously the quality of the final results of the data processing will be dependent on the richness of the raw data matrix! This richness is particularly influenced by the quality of the chosen series of model compounds and the diversity of the chromatographic packings and systems. Solutes will act as a probe to reveal particular characteristics of the packings and solvents. Reciprocally, packings and solvents will reveal the mechanism of retention responsible for solute behaviour.

For a general analysis, or for a more in-depth analysis of the raw data matrix, HAC and CFA are particularly valuable methods. HAC offers a global view between similarities of the considered elements and their filiation. It can be considered as a one-dimensional method of representing the filiation of the successive closest distances. In contrast, CFA exploits relative distances, in a hyperspace, to give the best simplified representation of this space with the choice of the successive main factors of inertia.

With both methods, the most widespread experimental field is the best to delineate the main trends in the chromatographic processes, *i.e.*, the chromatographic mode. Simultaneous analysis of chromatographic systems and solutes may determine the most significant subsets of the matrix. These subsets can offer more in-depth analysis of the information content specific to a chromatographic mode. Similarities and differences in chromatographic systems and solutes can be easily established. For example, study of packings has shown the interest of new packings such as AP, DNAP, DNB and TCP. These packings have strong specific interactions with solutes in the normal-phase mode. They are not redundant with others packings such as diol or NH_2. Similarity between columns packed with the same ODS bonded phase is not so obvious when a specific submatrix is selected. Solvent effects are analysed by selecting an appropriate subset. Classification of solvents can be obtained for any packings such as diol or ODS. Solute behaviour can be linked to physico-chemical parameters such as hydrophobicity or electronic effects.

HAC gives an overall view of similarities and successive filiation of packings, solvents and solutes. However, HAC of packings and solvents cannot be superposed on HAC of solutes as is possible with CFA. Chromatographic information, such as affinity or selectivity, is lost.

With CFA, systems and solutes are projected simultaneously on the same factorial plane. Proximities of packings or solutes permit the similarities between packings or solutes to be established. Moreover, proximities between systems and solutes measure the relative affinity and selectivity.

The flexibility of the considered data matrix must be emphasized. The data matrix can be extended to the raw chromatographic data of a newly available packing. This new packing can be characterized with respect to the currently large number of more or less standard packings.

REFERENCES

1 J. R. Chrétien, *Trends Anal. Chem.*, 6 (1987) 275.
2 L. R. Snyder, *J. Chromatogr.*, 16 (1978) 223.
3 R. Tijssen, H. A. H. Billiet and P. J. Schoenmakers, *J. Chromatogr.*, 122 (1976) 185.
4 D. L. Saunders, *Anal. Chem.*, 46 (1974) 470.
5 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *Chromatographia*, 15 (1982) 205.
6 L. R. Snyder, *Anal. Chem.*, 46 (1974) 1384.
7 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 185 (1979) 179.
8 P. J. Schoenmakers, H. A. H. Billiet and L. de Galan, *J. Chromatogr.*, 282 (1983) 107.
9 A. J. Hsu, R. J. Laub and S. J. Madden, *J. Liq. Chromatogr.*, 7 (1984) 615.
10 B. Walczak, M. Dreux, J. R. Chrétien, K. Szymoniak, M. Lafosse, L. Morin-Allory and J. P. Doucet, *J. Chromatogr.*, 353 (1986) 109.
11 B. Walczak, J. R. Chrétiens, M. Dreux, M. Lafosse, L. Morin-Allory, K. Szymoniak and F. Membrey, *J. Chromatogr.*, 353 (1986) 123.
12 B. Walczak, L. Morin-Allory, J. R. Chrétien, M. Lafosse and M. Dreux, *Chemometr. Intell. Lab. Syst.*, 1 (1986) 79.
13 B. Walczak, J. R. Chrétien, M. Dreux, L. Morin-Allory and M. Lafosse, *Chemometr. Intell. Lab. Syst.* 1 (1987) 177.
14 B. Walczak, M. Lafosse, J. R. Chrétien, M. Dreux and L. Morin-Allory, *J. Chromatogr.*, 369 (1986) 27.
15 J. R. Chrétien, B. Walczak, L. Morin-Allory, M. Dreux and M. Lafosse, *J. Chromatogr.*, 371 (1986) 253.
16 B. Walczak, L. Morin-Allory, M. Lafosse, M. Dreux and J. R. Chrétien, *J. Chromatogr.*, 395 (1987) 183.
17 B. Walczak, M. Dreux, J. R. Chrétien, L. Morin-Allory, M. Lafosse and G. Felix, *J. Chromatogr.*, 464 (1989) 237.
18 A. M. Siouffi, M. Righezza and G. Guiochon, *J. Chromatogr.*, 368 (1986) 189.